

WHAT IS CLAIMED IS:

- 1 1. A method for processing search results generated based
2 on a query, the method comprising:
 - 3 a) accepting the search results;
 - 4 b) accepting information derived from the query;
 - 5 c) generating a set of final search results from the
6 accepted search results based on the accepted
7 information.
- 1 2. The method of claim 1 wherein the act of generating a
2 set of final search results includes
 - 3 i) determining whether or not a candidate search
4 result is similar to a search result already in
5 the set of final search results; and
 - 6 ii) if it is determined that the candidate
7 search result is similar to a search result
8 already in the set of final search results, then
9 not adding the candidate search result to the set
10 of final search results.
- 1 3. The method of claim 2 wherein the act of determining
2 whether or not a candidate search result is similar to a
3 search result already in the set of final search results
4 includes
 - 5 A) extracting at least a part of the
6 candidate search result that is relevant to
7 the information derived from the query,
8 thereby generating first query-relevant
9 information;

10 B) extracting at least a part of the search
11 result already in the set of final search
12 results that is relevant to the information
13 derived from the query, thereby generating
14 second query-relevant information; and
15 C) determining whether or not the first
16 query-relevant information is similar to the
17 second query-relevant information,

18 wherein, if the first query-relevant information
19 is determined to be similar to the second query-relevant
20 information, then determining the candidate search result
21 to be similar to the search result already in the set of
22 final search results, and

23 wherein, if the first query-relevant information
24 is determined not to be similar to the second
25 query-relevant information, then determining the candidate
26 search result not to be similar to the search result
27 already in the set of final search results.

1 4. The method of claim 3 wherein the information derived
2 from the query includes query keywords, and

3 wherein the act of extracting at least a part of
4 the candidate search result that is relevant to the
5 information derived from the query, thereby generating
6 first query-relevant information includes:

7 1) defining a window as a first
8 predetermined number of characters;
9 2) applying the window to various
10 parts of a document corresponding to
11 the candidate search result;

12 3) for each of the various parts of
13 the document to which a window is
14 applied,
15 - determining the number of
16 keywords in the current part of
17 the document to determine a hit
18 count;
19 4) ranking the various parts of the
20 document to which a window is applied
21 based on its associated hit count; and
22 5) taking a second predetermined
23 number of the highest ranking various
24 parts of the document to define at
25 least a part of the first
26 query-relevant information.

1 5. The method of claim 3 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the candidate search result that is relevant to the
5 information derived from the query, thereby generating
6 first query-relevant information includes:
7 1) segmenting a document corresponding
8 to the candidate search result to
9 define a plurality of segments;
10 2) for each of the segments,
11 determining whether or not the segment
12 includes at least one of the query
13 keywords; and
14 3) for each of the segments, if it was
15 determined that the segment includes at
16 least one of the query keywords, then

17 adding the segment to the first
18 query-relevant information.

1 6. The method of claim 5 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 sentences.

1 7. The method of claim 5 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 paragraphs.

1 8. The method of claim 3 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the candidate search result that is relevant to the
5 information derived from the query, thereby generating
6 first query-relevant information includes:

- 7 1) segmenting a document corresponding
8 to the candidate search result to
9 define a plurality of segments;
- 10 2) for each of the segments,
11 determining whether or not the segment
12 includes at least a predetermined
13 number of the query keywords; and
- 14 3) for each of the segments, if it was
15 determined that the segment includes at
16 least the predetermined number of the
17 query keywords, then adding the segment
18 to the first query-relevant
19 information.

1 9. The method of claim 8 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 sentences.

1 10. The method of claim 8 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 paragraphs.

1 11. The method of claim 3 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the candidate search result that is relevant to the
5 information derived from the query, thereby generating
6 first query-relevant information includes:
7 1) segmenting a document corresponding
8 to the search result to define a
9 plurality of segments;
10 2) for each of the segments,
11 determining whether or not the segment
12 includes at a predetermined number of
13 different ones of the query keywords;
14 and
15 3) for each of the segments, if it was
16 determined that the segment includes at
17 least the predetermined number of
18 different ones of the query keywords,
19 then adding the segment to the first
20 query-relevant information.

1 12. The method of claim 11 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 sentences.

1 13. The method of claim 11 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 paragraphs.

1 14. The method of claim 1 wherein the act of generating a
2 set of final search results includes
3 i) determining whether or not a candidate search
4 result is similar to a search result already in
5 the set of final search results; and
6 ii) adding the search results to the set of
7 final search results only if it is determined
8 that the candidate search result is not similar
9 to any search results already in the set of final
10 search result.

1 15. The method of claim 14 wherein the act of determining
2 whether or not a candidate search result is similar to a
3 search result already in the set of final search results
4 includes
5 A) extracting at least a part of the
6 candidate search result that is relevant to
7 the information derived from the query,
8 thereby generating first query-relevant
9 information;
10 B) extracting at least a part of the search
11 result already in the set of final search

12 results that is relevant to the information
13 derived from the query, thereby generating
14 second query-relevant information; and
15 C) determining whether or not the first
16 query-relevant information is similar to the
17 second query-relevant information,
18 wherein, if the first query-relevant information
19 is determined to be similar to the second query-relevant
20 information, then determining the candidate search result
21 to be similar to the search result already in the set of
22 final search results, and
23 wherein, if the first query-relevant information
24 is determined not to be similar to the second
25 query-relevant information, then determining the candidate
26 search result not to be similar to the search result
27 already in the set of final search results.

1 16. The method of claim 15 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the candidate search result that is relevant to the
5 information derived from the query, thereby generating
6 first query-relevant information includes:
7 1) defining a window as a first
8 predetermined number of characters;
9 2) applying the window to various
10 parts of a document corresponding to
11 the candidate search result;
12 3) for each of the various parts of
13 the document to which a window is
14 applied,

15 - determining the number of
16 keywords in the current part of
17 the document to determine a hit
18 count;
19 4) ranking the various parts of the
20 document to which a window is applied
21 based on its associated hit count; and
22 5) taking a second predetermined
23 number of the highest ranking various
24 parts of the document to define at
25 least a part of the first
26 query-relevant information.

1 17. The method of claim 15 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the candidate search result that is relevant to the
5 information derived from the query, thereby generating
6 first query-relevant information includes:
7 1) segmenting a document corresponding
8 to the search result to define a
9 plurality of segments;
10 2) for each of the segments,
11 determining whether or not the segment
12 includes at least one of the query
13 keywords; and
14 3) for each of the segments, if it was
15 determined that the segment includes at
16 least one of the query keywords, then
17 adding the segment to the first
18 query-relevant information.

1 18. The method of claim 17 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 sentences.

1 19. The method of claim 17 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 paragraphs.

1 20. The method of claim 15 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the candidate search result that is relevant to the
5 information derived from the query, thereby generating
6 first query-relevant information includes:
7 1) segmenting a document corresponding
8 to the candidate search result to
9 define a plurality of segments;
10 2) for each of the segments,
11 determining whether or not the segment
12 includes at a predetermined number of
13 the query keywords; and
14 3) for each of the segments, if it was
15 determined that the segment includes at
16 least the predetermined number of the
17 query keywords, then adding the segment
18 to the first query-relevant
19 information.

1 21. The method of claim 20 wherein the act of segmenting a
2 document corresponding to the candidate search result to

3 define a plurality of segments, segments the document into
4 sentences.

1 22. The method of claim 20 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 paragraphs.

1 23. The method of claim 15 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the candidate search result that is relevant to the
5 information derived from the query, thereby generating
6 first query-relevant information includes:

- 7 1) segmenting a document corresponding
8 to the candidate search result to
9 define a plurality of segments;
- 10 2) for each of the segments,
11 determining whether or not the segment
12 includes at a predetermined number of
13 different ones of the query keywords;
14 and
- 15 3) for each of the segments, if it was
16 determined that the segment includes at
17 least the predetermined number of
18 different ones of the query keywords,
19 then adding the segment to the first
20 query-relevant information.

1 24. The method of claim 23 wherein the act of segmenting a
2 document corresponding to the candidate search result to

3 define a plurality of segments, segments the document into
4 sentences.

1 25. The method of claim 23 wherein the act of segmenting a
2 document corresponding to the candidate search result to
3 define a plurality of segments, segments the document into
4 paragraphs.

1 26. A method for determining whether or not a first
2 document corresponding to a first search result is similar
3 to a second document corresponding to a second search
4 result, the method comprising:

- 5 a) accepting a query that generated the first and
6 second search results;
 - 7 b) extracting at least a part of the first document
8 that is relevant to the information derived from the
9 query, thereby generating first query-relevant
10 information;
 - 11 c) extracting at least a part of the second document
12 that is relevant to the information derived from the
13 query, thereby generating second query-relevant
14 information; and
 - 15 d) determining whether or not the first
16 query-relevant information is similar to the second
17 query-relevant information,
- 18 wherein, if the first query-relevant information
19 is determined to be similar to the second query-relevant
20 information, then determining the first document to be
21 similar to the second document, and
- 22 wherein, if the first query-relevant information
23 is determined not to be similar to the second

24 query-relevant information, then determining the first
25 document not to be similar to the second document.

1 27. The method of claim 26 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the first document that is relevant to the information
5 derived from the query, thereby generating first
6 query-relevant information includes:

- 7 1) defining a window as a first
8 predetermined number of characters;
- 9 2) applying the window to various
10 parts of the first document;
- 11 3) for each of the various parts of
12 the first document to which a window is
13 applied,
 - 14 - determining the number of
15 keywords in the current part of
16 the first document to determine a
17 hit count;
- 18 4) ranking the various parts of the
19 first document to which a window is
20 applied based on its associated hit
21 count; and
- 22 5) taking a second predetermined
23 number of the highest ranking various
24 parts of the first document to define
25 at least a part of the first
26 query-relevant information.

1 28. The method of claim 26 wherein the information derived
2 from the query includes query keywords, and

3 wherein the act of extracting at least a part of
4 the first documents that is relevant to the information
5 derived from the query, thereby generating first
6 query-relevant information includes:

- 7 1) segmenting the first document to
8 define a plurality of segments;
- 9 2) for each of the segments,
10 determining whether or not the segment
11 includes at least one of the query
12 keywords; and
- 13 3) for each of the segments, if it was
14 determined that the segment includes at
15 least one of the query keywords, then
16 adding the segment to the first
17 query-relevant information.

1 29. The method of claim 28 wherein the act of segmenting
2 the first document to define a plurality of segments,
3 segments the first document into sentences.

1 30. The method of claim 28 wherein the act of segmenting
2 the first document to define a plurality of segments,
3 segments the first document into paragraphs.

1 31. The method of claim 26 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the first document that is relevant to the information
5 derived from the query, thereby generating first
6 query-relevant information includes:

- 7 1) segmenting the first document to
8 define a plurality of segments;

9 2) for each of the segments,
10 determining whether or not the segment
11 includes at a predetermined number of
12 the query keywords; and
13 3) for each of the segments, if it was
14 determined that the segment includes at
15 least the predetermined number of the
16 query keywords, then adding the segment
17 to the first query-relevant
18 information.

1 32. The method of claim 31 wherein the act of segmenting
2 the first document to define a plurality of segments,
3 segments the first document into sentences.

1 33. The method of claim 31 wherein the act of segmenting
2 the first document to define a plurality of segments,
3 segments the first document into paragraphs.

1 34. The method of claim 26 wherein the information derived
2 from the query includes query keywords, and
3 wherein the act of extracting at least a part of
4 the first document that is relevant to the information
5 derived from the query, thereby generating first
6 query-relevant information includes:
7 1) segmenting the first document to
8 define a plurality of segments;
9 2) for each of the segments,
10 determining whether or not the segment
11 includes at a predetermined number of
12 different ones of the query keywords;
13 and

14 3) for each of the segments, if it was
15 determined that the segment includes at
16 least the predetermined number of
17 different ones of the query keywords,
18 then adding the segment to the first
19 query-relevant information.

1 35. The method of claim 34 wherein the act of segmenting
2 the first document to define a plurality of segments,
3 segments the document into sentences.

1 36. The method of claim 34 wherein the act of segmenting
2 the first document to define a plurality of segments,
3 segments the document into paragraphs.

1 37. A machine-readable medium including machine executable
2 instructions which, when executed by a machine, processes
3 search results generated based on a query by:

- 4 a) accepting the search results;
- 5 b) accepting information derived from the query;
- 6 c) generating a set of final search results from the
- 7 accepted search results based on the accepted
- 8 information.

1 38. The machine-readable medium of claim 37 including
2 further machine executable instructions which, when
3 executed by a machine, generate the set of final search
4 results by

- 5 i) determining whether or not a candidate search
- 6 result is similar to a search result already in
- 7 the set of final search results; and

8 ii) adding the candidate search result to the
9 set of final search results only if it is
10 determined that the candidate search result is
11 not similar to any search results already in the
12 set of final search result.

1 39. The machine-readable medium of claim 38 including
2 further machine executable instructions which, when
3 executed by a machine determine whether or not a search
4 result is similar to a search result already in the set of
5 final search results by
6 A) extracting at least a part of the search
7 result that is relevant to the information
8 derived from the query, thereby generating
9 first query-relevant information;
10 B) extracting at least a part of the search
11 result already in the set of final search
12 results that is relevant to the information
13 derived from the query, thereby generating
14 second query-relevant information; and
15 C) determining whether or not the first
16 query-relevant information is similar to the
17 second query-relevant information,
18 wherein, if the first query-relevant information
19 is determined to be similar to the second query-relevant
20 information, then determining the search results to be
21 similar to the search result already in the set of final
22 search results, and
23 wherein, if the first query-relevant information
24 is determined not to be similar to the second
25 query-relevant information, then determining the search

26 results not to be similar to the search result already in
27 the set of final search results.

1 40. An apparatus for processing search results generated
2 based on a query, the apparatus comprising:
3 a) a storage facility for storing the search results
4 and for storing information derived from the query;
5 and
6 b) a final results generator for generating a set of
7 final search results from the search results stored in
8 the storage facility based on the information stored
9 in the storage facility.

1 41. The apparatus of claim 40 wherein the final set
2 generator includes
3 i) a similarity determination facility for
4 determining whether or not a search result is
5 similar to a search result already in the set of
6 final search results; and
7 ii) means for adding the search results to the
8 set of final search results only if the
9 similarity determination facility determines that
10 the search result is not similar to any search
11 results already in the set of final search
12 result.

1 42. The method of claim 41 wherein the similarity
2 determination facility includes
3 A) a query-relevant extraction facility for
4 extracting at least a part of the search
5 result that is relevant to the information
6 derived from the query, thereby generating

7 first query-relevant information, and for
8 extracting at least a part of the search
9 result already in the set of final search
10 results that is relevant to the information
11 derived from the query, thereby generating
12 second query-relevant information; and
13 B) a query-relevant similarity
14 determination facility for determining
15 whether or not the first query-relevant
16 information is similar to the second
17 query-relevant information,
18 wherein, if the query-relevant similarity
19 determination facility determines that the first
20 query-relevant information is similar to the second
21 query-relevant information, then the similarity
22 determination facility determines the search result to be
23 similar to the search result already in the set of final
24 search results, and
25 wherein, if the query-relevant similarity
26 determination facility determines that the first
27 query-relevant information is not similar to the second
28 query-relevant information, then the similarity
29 determination facility determines the search result not to
30 be similar to the search result already in the set of final
31 search results.

1 43. A storage facility including at least one
2 machine-readable medium storing information comprising:
3 a) ranked query results;
4 b) query-relevant parts of documents corresponding to
5 the ranked query results; and
6 c) a final set of query results,

7 wherein the final set of query results is a
8 sub-set of the ranked query results, and
9 wherein the final set of query results does not
10 include any two query results corresponding to documents
11 that have similar query-relevant parts.

1 44. The storage facility of claim 43, the stored
2 information further comprising:

3 d) documents corresponding to the ranked query
4 results.

1 45. A method for processing search results generated based
2 on a query, the method comprising:

3 a) accepting the search results;
4 b) accepting information derived from the query;
5 c) accepting documents associated with the search
6 results;
7 d) extracting portions of the documents associated
8 with the search results based on the information
9 derived from the query to generate query-relevant
10 information for each of the documents; and
11 e) generating a set of final search results from the
12 accepted search results based on the query-relevant
13 information.